FIG. 1

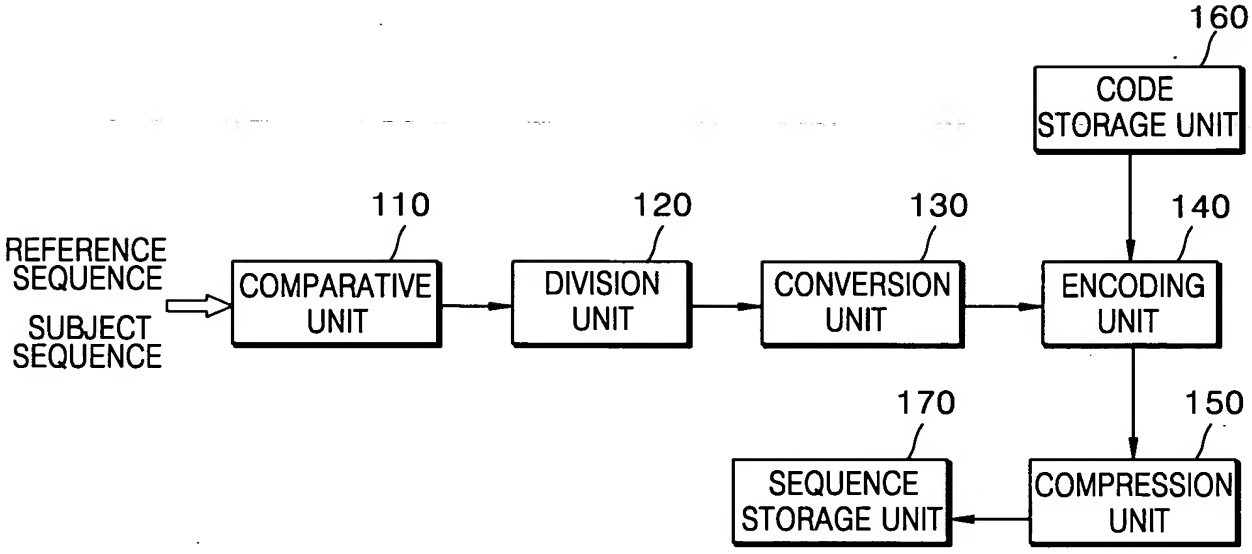
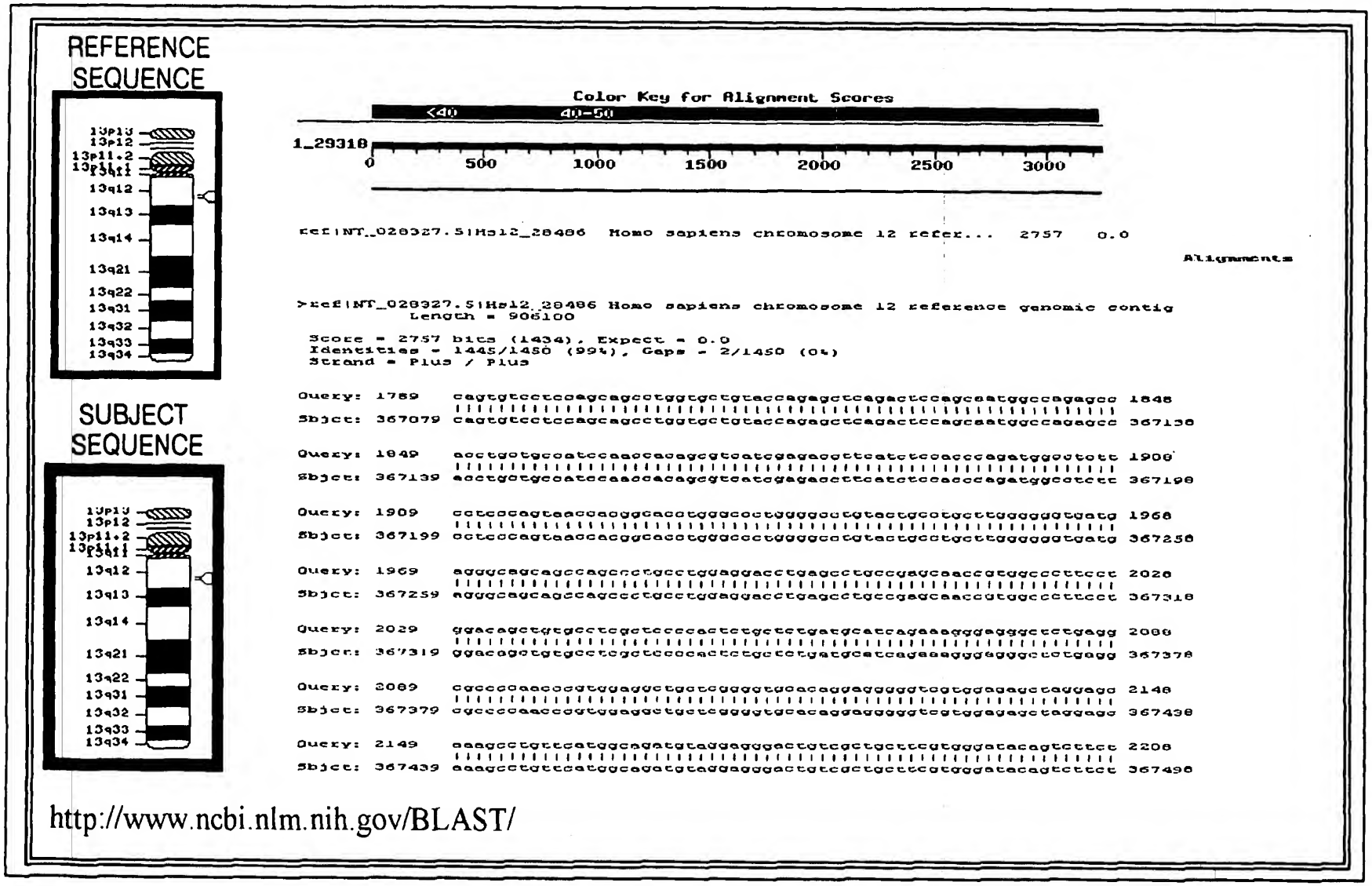

160
CODE STORAGE UNIT
110
120
130
140
REFERENCE SEQUENCE
SUBJECT SEQUENCE
COMPARATIVE UNIT
DIVISION UNIT
CONVERSION UNIT
ENCODING UNIT
170
150
SEQUENCE STORAGE UNIT
COMPRESSION UNIT

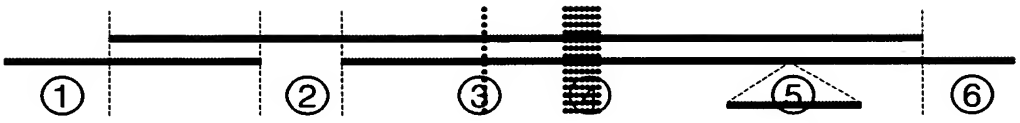
FIG. 2



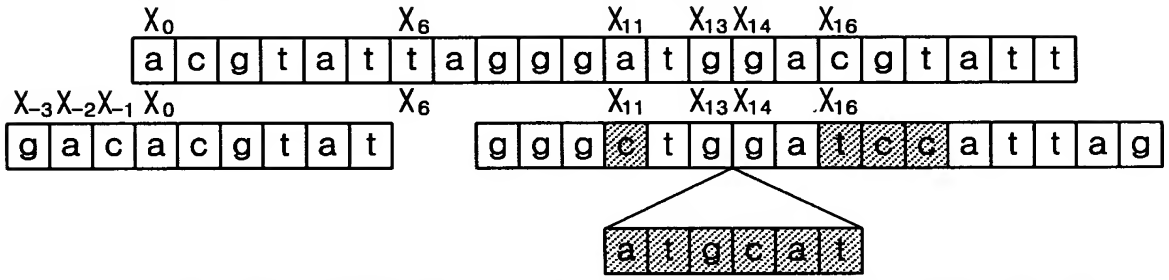
http://www.ncbi.nlm.nih.gov/BLAST/

BEST AVAILABLE COPY

FIG. 3



① Start region mismatch
② Blank region
③ Single base pair mismatch
④ Multiple base pair mismatch
⑤ Inserted region
⑥ End region mismatch



/-3~3gac/3/6/2/3~1c/1/1~6atgcat/1/2~3tcc/3/4~2ag/2

FIG. 4

4BITS CODES FOR ENCODING INFORMATIN ABOUT THE DIFFERENCE BETWEEN DNA SEQUENCES	
0 : 0000	8 : 1000
1 : 0001	9 : 1001
2 : 0010	A : 1010
3 : 0011	T : 1011
4 : 0100	G : 1100
5 : 0101	C : 1101
6 : 0110	/ : 1110
7 : 0111	~ : 1111

FIG. 5

MODY3 exons (5552 characters = 5552 byte)
ref|NT_028327.5|Hs12_28486 Homo sapiens chromosome 12 reference genomic contig
Length = 906100

Exon2 /354759-1a/1	Exon789 /363265-1t/1	Exon789 /363384-1t/1
Exon3 /359197-1t/1	Exon789 /363261-1t/1	Exon789 /363391-1t/1
Exon456 /362119-1a/1	Exon789 /363267-1t/1	Exon789 /363394-1t/1
Exon456 /362178-1a/1	Exon789 /363273-1t/1	Exon789 /363641-1a/1
Exon789 /363208-1t/1	Exon789 /363278-1t/1	Exon789 /363689-1a/1

/354759-1a/1/4438-1a/1/2922-1a/1/59-1a/1/30-1a/53-1t/1/4-1t/1/2-1/t/1/6-1t/1/5-1t/1/6
1t/1/7-1t/1/3-1t/1/247-1a/1/248-1a/1

122byte

11100011010101000111010110011111000011010111000011110010001000011100011100001101011100001111000101001001000101110000110101110000111100101100111100001101011100001111000110000111000011010111001010011111000011011111000011100001110110111000011110000111001101110000111011111000011110000111000011010111100001111000111011100001110000111001101110000110111110000111100001111100001101111100001111000110100111000011101111000011101000100011111000011010111000011110001001001000111000011010111000011110001001001000111000011010111000011110001001000111000011010111000011110001001001000111000011010111000011110001001000111000011010111000011110001001001000111000011010111000011110001001001000111000011010111000011110001

FIG. 6

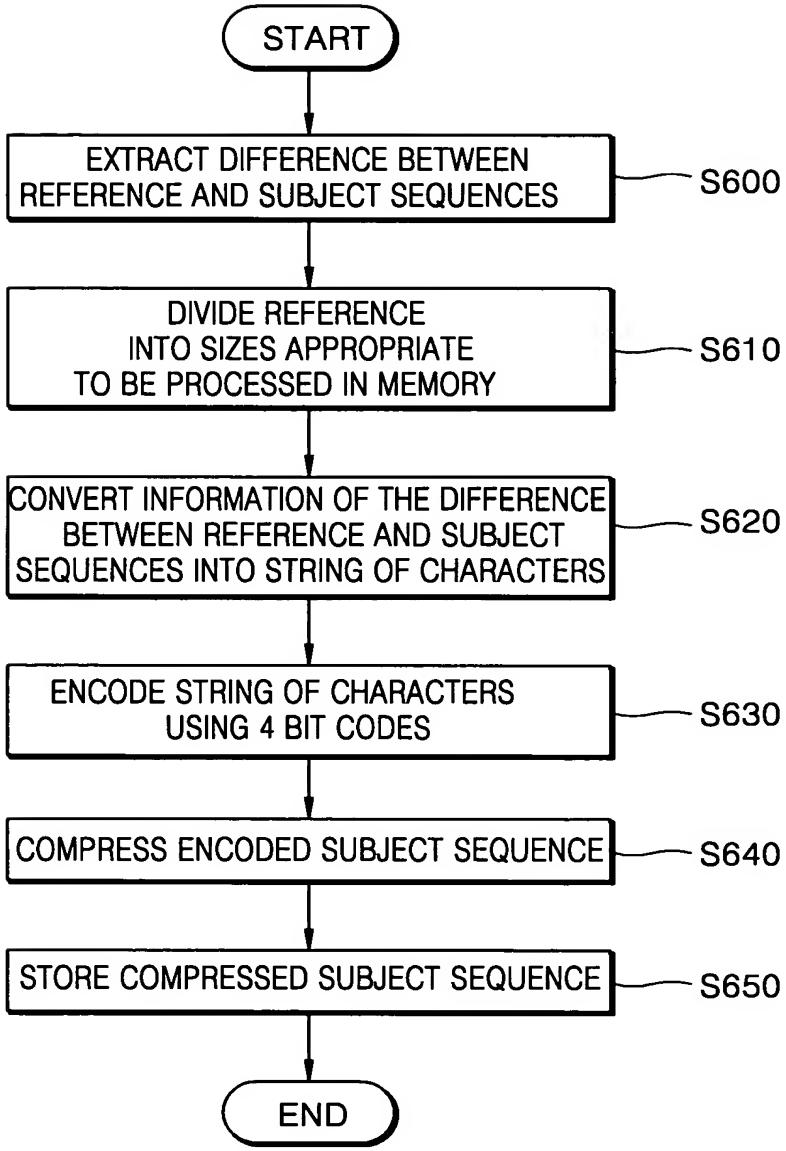

START
EXTRACT DIFFERENCE BETWEEN REFERENCE AND SUBJECT SEQUENCES
S600
DIVIDE REFERENCE INTO SIZES APPROPRIATE TO BE PROCESSED IN MEMORY
S610
CONVERT INFORMATION OF THE DIFFERENCE BETWEEN REFERENCE AND SUBJECT SEQUENCES INTO STRING OF CHARACTERS
S620
ENCODE STRING OF CHARACTERS USING 4 BIT CODES
S630
COMPRESS ENCODED SUBJECT SEQUENCE
S640
STORE COMPRESSED SUBJECT SEQUENCE
S650
END

FIG. 7

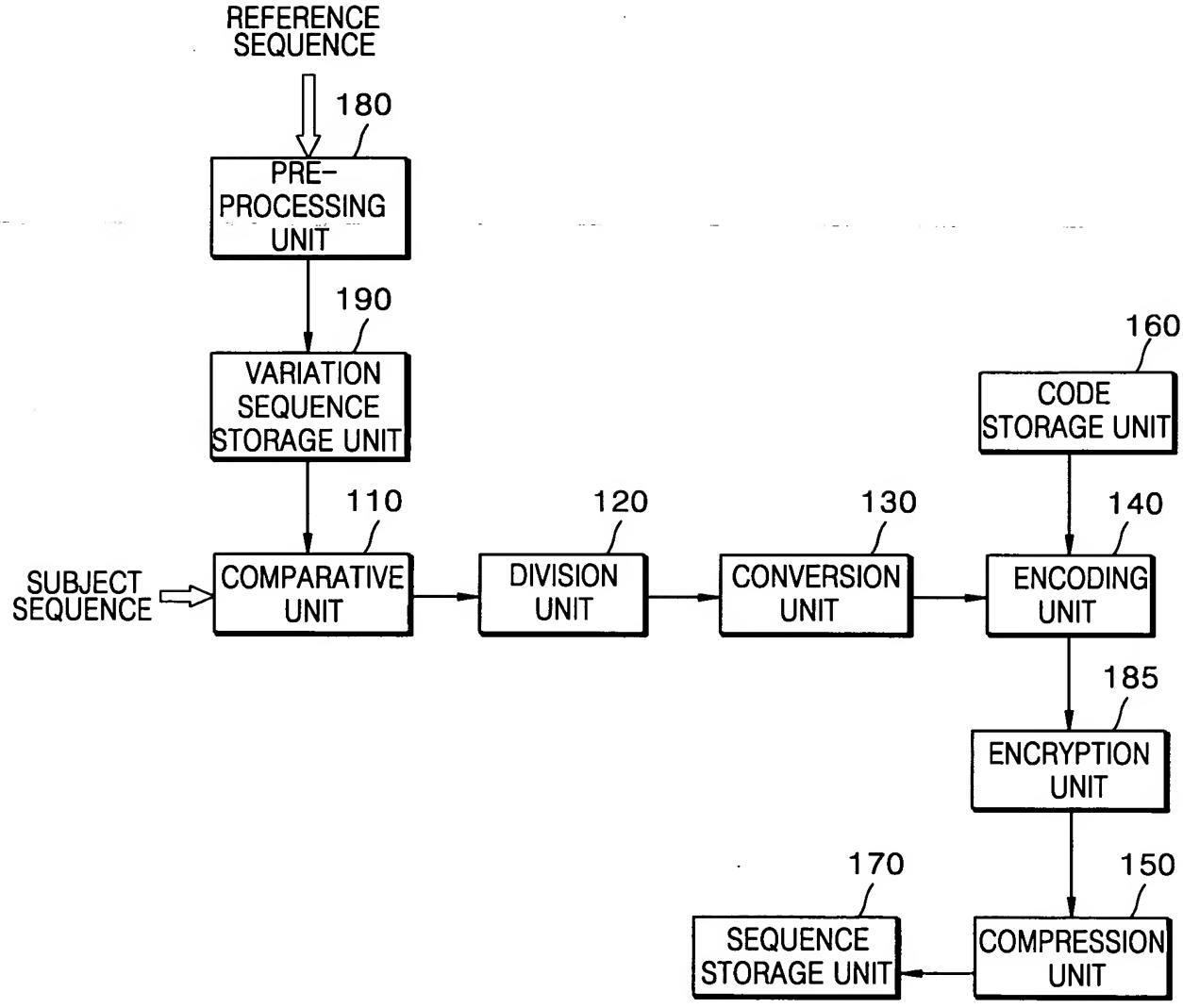

REFERENCE SEQUENCE
180
PRE-PROCESSING UNIT
190
VARIATION SEQUENCE STORAGE UNIT
160
CODE STORAGE UNIT
110
120
130
140
SUBJECT SEQUENCE
COMPARATIVE UNIT
DIVISION UNIT
CONVERSION UNIT
ENCODING UNIT
185
ENCRYPTION UNIT
170
150
SEQUENCE STORAGE UNIT
COMPRESSION UNIT

FIG. 8

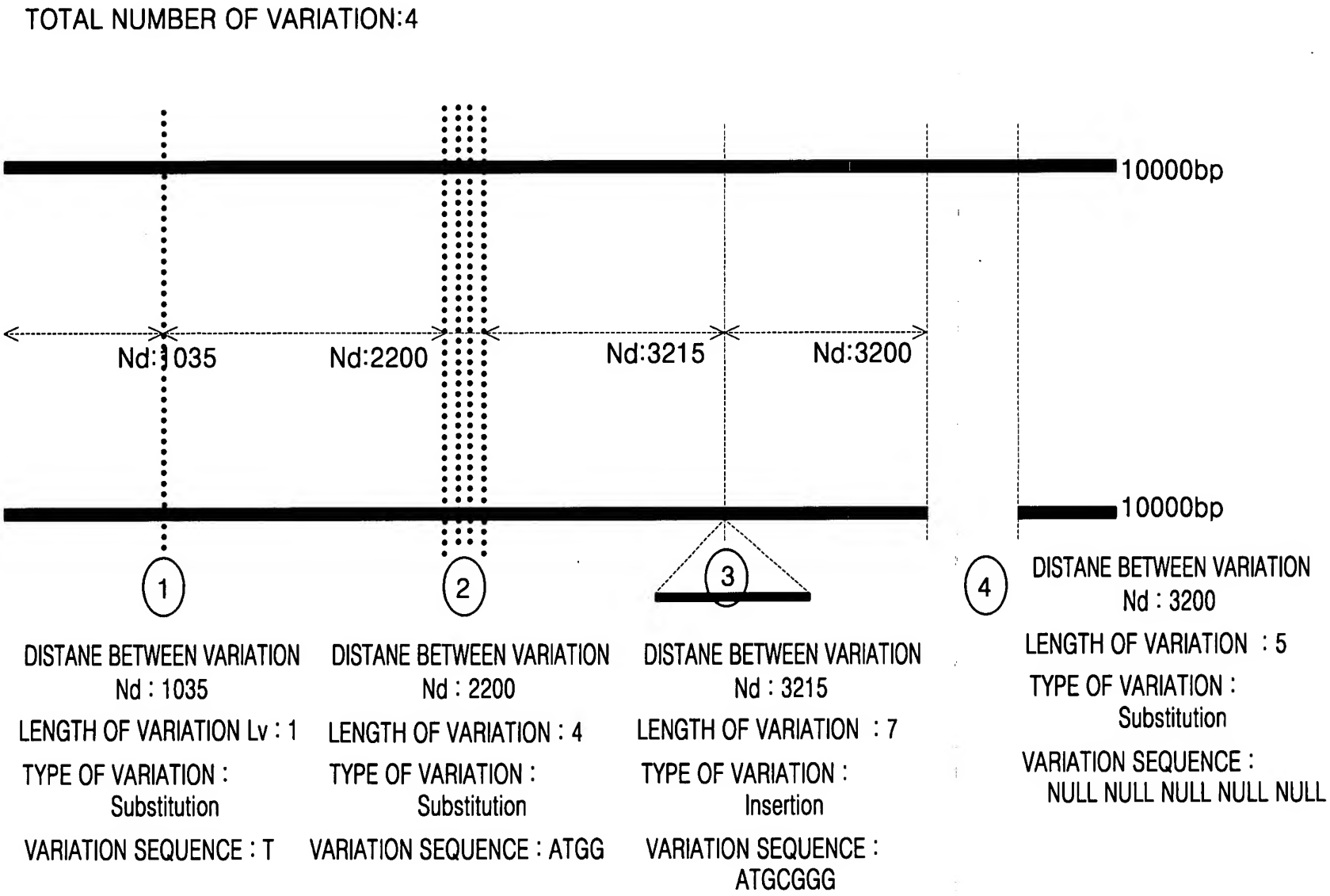
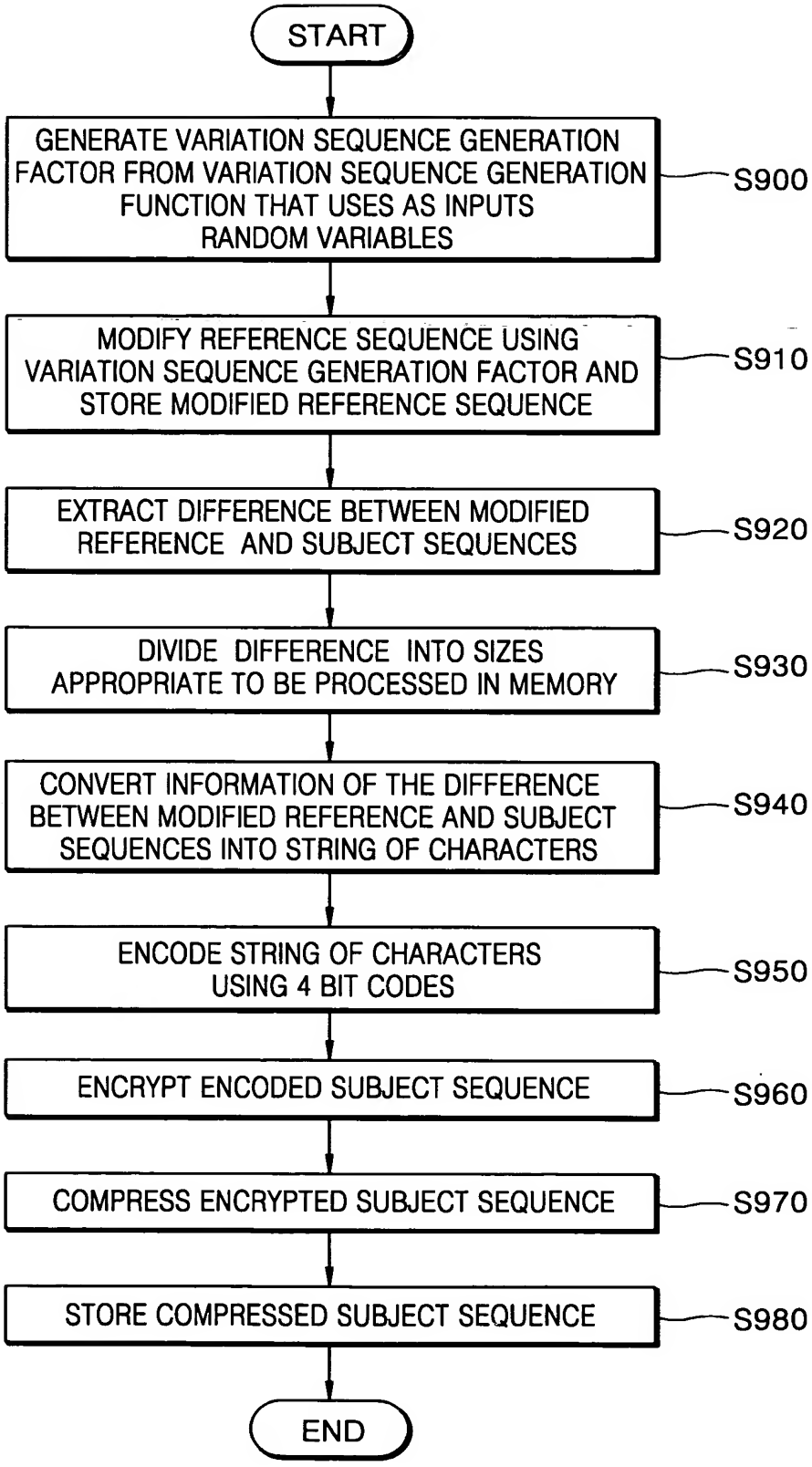

TOTAL NUMBER OF VARIATION:4
10000bp
Nd:1035
Nd:2200
Nd:3215
Nd:3200
10000bp
1
2
3
4
DISTANE BETWEEN VARIATION
Nd : 1035
LENGTH OF VARIATION Lv : 1
TYPE OF VARIATION :
Substitution
VARIATION SEQUENCE : T
DISTANE BETWEEN VARIATION
Nd : 2200
LENGTH OF VARIATION : 4
TYPE OF VARIATION :
Substitution
VARIATION SEQUENCE : ATGG
DISTANE BETWEEN VARIATION
Nd : 3215
LENGTH OF VARIATION : 7
TYPE OF VARIATION :
Insertion
VARIATION SEQUENCE :
ATGCGGG
DISTANE BETWEEN VARIATION
Nd : 3200
LENGTH OF VARIATION : 5
TYPE OF VARIATION :
Substitution
VARIATION SEQUENCE :
NULL NULL NULL NULL NULL

FIG. 9


START
GENERATE VARIATION SEQUENCE GENERATION FACTOR FROM VARIATION SEQUENCE GENERATION FUNCTION THAT USES AS INPUTS RANDOM VARIABLES
S900
MODIFY REFERENCE SEQUENCE USING VARIATION SEQUENCE GENERATION FACTOR AND STORE MODIFIED REFERENCE SEQUENCE
S910
EXTRACT DIFFERENCE BETWEEN MODIFIED REFERENCE AND SUBJECT SEQUENCES
S920
DIVIDE DIFFERENCE INTO SIZES APPROPRIATE TO BE PROCESSED IN MEMORY
S930
CONVERT INFORMATION OF THE DIFFERENCE BETWEEN MODIFIED REFERENCE AND SUBJECT SEQUENCES INTO STRING OF CHARACTERS
S940
ENCODE STRING OF CHARACTERS USING 4 BIT CODES
S950
ENCRYPT ENCODED SUBJECT SEQUENCE
S960
COMPRESS ENCRYPTED SUBJECT SEQUENCE
S970
STORE COMPRESSED SUBJECT SEQUENCE
S980
END